Consistency of Concave Regression with an Application to Current-Status Data

Lutz Dümbgen, University of Bern
Sandra Freitag, University of Kiel
Geurt Jongbloed, Vrije Universiteit Amsterdam

May 2002 (revised March 2004)
Appeared in *Mathematical Methods of Statistics* ***13***, pp. 69-81.

Abstract. We consider the problem of nonparametric estimation of a concave regression function F. We show that the supremum distance between the least squares estimator and F on a compact interval is typically of order $(\log(n)/n)^{2/5}$. This entails rates of convergence for the estimator's derivative. Moreover, we discuss the impact of additional constraints on F such as monotonicity and pointwise bounds. Then we apply these results to the analysis of current status data, where the distribution function of the event times is assumed to be concave.

AMS 1991 subject classifications: primary 62G05; secondary 62E20

Key words: interval censoring, least squares, monotone density, subgaussian errors, supremum norm

Work supported by Deutsche Forschungsgemeinschaft

1 Introduction

Suppose that we observe $(t_1, Y_1), (t_2, Y_2), \ldots, (t_n, Y_n)$ with fixed numbers $t_1 \leq t_2 \leq \cdots \leq t_n$ and independent random variables $Y_1, Y_2, \ldots, Y_n$. Let

$$\mathbb{E}(Y_i) \;=\; F(t_i) \tag{1}$$

for some unknown regression function $F : \boldsymbol{J} \to \boldsymbol{R}$, where $\boldsymbol{J}$ is some interval containing the design points t_i. In various applications, e.g. in econometrics, the regression function F is known to be concave. Then it is possible to estimate F without further assumptions by the method of least squares. That means, let $\widehat{F}$ be any concave function on $\boldsymbol{J}$

minimizing

$$q(G) \;:=\; \sum_{i=1}^{n} (Y_i - G(t_i))^2$$

over the set of all concave functions G on $\boldsymbol{J}$. This estimator $\widehat{F}$ exists and is uniquely determined on the set $\{t_1, t_2, \ldots, t_n\}$, because $q(G)$ is strictly convex and coercive in $(G(t_i))_{i=1}^n$. Basic properties of $\widehat{F}$ and various consistency results have been derived. Hanson and Pledger (1976) prove uniform consistency, whereas Mammen (1991) and Groeneboom et al. (2001) concentrate on pointwise limit theorems. The present paper focuses on the supremum norm of $\widehat{F} - F$ and its derivative. Section 2 contains the main results. In particular, under certain regularity assumptions, the supremum norm of $\widehat{F} - F$ on a bounded interval is of stochastic order $(\log(n)/n)^{2/5}$, while $\widehat{F}' - F'$ converges at rate $(\log(n)/n)^{1/5}$.

The impact of additional constraints such as isotonicity and pointwise bounds is discussed in Section 3. In Section 4 we consider the special case of current-status data. Here $t_1, \ldots, t_n \in (0, \infty)$ are inspection times, and

$$Y_i \;=\; 1\{X_i \le t_i\}$$

with independent event times $X_1, X_2, \ldots, X_n$ having distribution function F on $[0, \infty]$. If F is assumed to be concave on $[0, \infty)$, then (1) holds, and the main results from Sections 2 and 3 carry over to the least-squares estimator $\widehat{F}$ for the distribution function F.

All proofs are deferred to Section 5. Note that the techniques developed here are different from the entropy-based approach of van de Geer (2000). While she uses covering numbers for the set of all potential regression functions, we are using a much smaller class of “caricatures” (piecewise linear functions) for the difference between true and estimated curve.

2 Uniform consistency

We consider a triangular scheme of observations $t_i = t_{n,i}$ and $Y_i = Y_{n,i}$ but suppress the additional subscript n for notational simplicity. Let M_n be the empirical distribution of the design points t_i, i.e.

$$M_n(B) \;:=\; n^{-1} \sum_{i=1}^{n} 1\{t_i \in B\}$$

for $B \subset \boldsymbol{R}$. In this section, we analyze the asymptotic behavior of $\widehat{F} = \widehat{F}_n$ on a fixed compact interval $[a, b] \subset \boldsymbol{J}$ under certain conditions on M_n and the errors

$$E_i = E_{n,i} \ := \ Y_i - F(t_i).$$

Condition I. There is a constant $C > 0$ such that

$$\liminf_{n\to\infty} \frac{M_n[a_n, b_n]}{b_n - a_n} \ \geq \ C$$

whenever $a \leq a_n < b_n \leq b$ such that $\liminf_{n\to\infty} n^{1/3}(b_n - a_n) > 0$.

Condition II. For some constant $\sigma > 0$,

$$\max_{i=1,\ldots,n} \mathbb{E}\exp(\lambda E_i) \ \leq \ \exp(\sigma^2\lambda^2/2) \quad \text{for all } \lambda \in \boldsymbol{R}.$$

Condition III. There are constants $\beta \in [1, 2]$ and L such that for arbitrary $s, t \in [a, b]$,

$$\begin{cases} |F(s) - F(t)| \leq L|s-t| & \text{if } \beta = 1, \\ |F'(s) - F'(t)| \leq L|s-t|^{\beta-1} & \text{if } \beta > 1. \end{cases}$$

If, for instance, $t_1, \ldots, t_n$ are the order statistics of independent random variables $T_1, \ldots, T_n$ with distribution Q satisfying $Q[a', b'] \geq C(b' - a')$ for $a \leq a' < b' \leq b$, then Condition I is satisfied almost surely. It is also satisfied with $[a, b] = [0, 1]$ in case of regular design points $t_i = i/n$.

Condition II is satisfied with $\lambda = 2$ if, for instance, the errors E_i are Gaussian with standard deviation not greater than σ.

Condition III is always satisfied with $\beta = 1$ and some L, provided that $a > \inf(\boldsymbol{J})$ and $b < \sup(\boldsymbol{J})$. If F is twice differentiable with $|F''| \leq L$, then Condition III holds with $\beta = 2$.

Theorem 1 *Suppose that Conditions I-III are satisfied. Then*

$$\begin{aligned} \max_{t\in[a,b]} (\widehat{F} - F)(t) &= O_p\left(\rho_n^{\beta/(2\beta+1)}\right), \\ \max_{t\in[a+\delta_n, b-\delta_n]} (F - \widehat{F})(t) &= O_p\left(\rho_n^{\beta/(2\beta+1)}\right), \end{aligned}$$

where $\rho_n := \log(n)/n$ *and* $\delta_n := \rho_n^{1/(2\beta+1)}$.

As a direct consequence of this theorem we get a result about uniform consistency of the derivative $\widehat{F}'$ in the case $\beta > 1$.

Corollary 1 *Suppose that Conditions I-III are satisfied with $\beta > 1$. Then*

$$\max_{t \in [a+\delta_n, b-\delta_n]} |\widehat{F}'(t) - F'(t)| = O_p\left(\rho_n^{(\beta-1)/(2\beta+1)}\right), \tag{2}$$

where $\widehat{F}'$ can be interpreted either as left- or rightsided derivative.

Groeneboom et al. (2001, Theorem 6.3) establish the pointwise limit behavior of the least squares estimator, featuring a rate of $n^{-2/5}$ for the concave function $\widehat{F}$ and a rate of $n^{-1/5}$ for its derivative $\widehat{F}'$ at a fixed point. These rates are established under a smoothness assumption which corresponds to the current situation with $\beta = 2$.

The rates derived here are indeed optimal. More precisely, in the special case of gaussian errors E_i with variance σ^2 and equidistant design points $t_i = i/n$ one can modify the arguments of Ibragimov and Khasminskii (1980) in order to show that for any nondegenerate interval $[a, b] \subset [0, 1]$ and parameters $\beta \in [0, 1]$, $L > 0$, there exist strictly positive constants $c(\beta, L)$ and $c'(\beta, L)$ such that

$$\inf_{\widehat{F}} \sup_{F \in \mathcal{F}_{\mathrm{conc}}(\beta, L)} \mathbb{P}\Big\{\max_{t \in [a,b]} |\widehat{F}(t) - F(t)| \geq c(\beta, L)\rho_n^{\beta/(2\beta+1)}\Big\} \ \to \ 1$$

and

$$\inf_{\widehat{F}} \sup_{F \in \mathcal{F}_{\mathrm{conc}}(\beta, L)} \mathbb{P}\Big\{\max_{t \in [a,b]} |\widehat{F}'(t) - F'(t)| \geq c'(\beta, L)\rho_n^{(\beta-1)/(2\beta+1)}\Big\} \ \to \ 1$$

as $n \to \infty$. Here $\mathcal{F}_{\mathrm{conc}}(\beta, L)$ stands for the set of all concave functions satisfying Condition III.

3 Additional constraints

In several settings the regression function F is assumed to satisfy additional constraints such as isotonicity or certain pointwise bounds. Then it is natural to impose the same additional restrictions on the estimator $\widehat{F}$. Intuitively one would expect that this improves the estimator, but there seems to be no simple argument for this claim. In terms of rates of convergence there is no improvement: The minimax results mentioned at the end of Section 2 remain valid if the function F is assumed in addition to be isotonic and to satisfy finitely many inequalities of the type $c_o \leq F(s_o) \leq d_o$.

Let $\mathcal{F}_{(1)}$ be the set of all concave and *isotonic* functions on $\boldsymbol{J}$. Furthermore, let $\mathcal{F}_{(2)}$ be the set of all concave functions G on $\boldsymbol{J}$ satisfying the inequalities

$$v_i \ \le \ G(s_i) \ \le \ w_i \quad \text{for } 1 \le i \le I$$

for a finite number I of points $s_i \in \boldsymbol{J}$ and numbers $-\infty \le v_i \le w_i \le \infty$. Finally, let $\mathcal{F}_{(3)} := \mathcal{F}_{(1)} \cap \mathcal{F}_{(2)}$. Then we define the restricted LS estimators

$$\widehat{F}_{(j)} \ := \ \operatorname*{arg\,min}_{G \in \mathcal{F}_{(j)}} q(G),$$

assuming tacitly that the set $\mathcal{F}_{(j)}$ is nonvoid.

Theorem 2 *For a given $j \in \{1,2,3\}$, suppose that $F \in \mathcal{F}_{(j)}$, and let Conditions I–III be satisfied. Then the conclusions of Theorem 1 and Corollary 1 remain true for $\widehat{F}_{(j)}$ in place of $\widehat{F}$.*

4 Current status data

A special example for the present setting is the current status model. The basic object of interest is a distribution function F on $[0,\infty]$ modelling a random event time, e.g. the time of onset of a certain disease. Suppose that $X_1, X_2, \ldots$ are event times with distribution function F, but we are not able to observe these directly. Instead, given inspection time points $0 < t_1 \le t_2 \le \cdots \le t_n < \infty$, we observe $Y_i = 1\{X_i \le t_i\}$ for $1 \le i \le n$.

The standard current status model and estimators for the distribution function based on such data are understood well by now; see for instance Groeneboom and Wellner (1992). An intensely studied estimator for F is the nonparametric maximum likelihood estimator (NPMLE) which maximizes

$$\ell(G) \ = \ \sum_{i=1}^{n} \left[Y_i \log G(t_i) + (1 - Y_i) \log(1 - G(t_i)) \right]$$

over the class of all distribution functions G on $[0,\infty]$. This estimator may be chosen to be a step function with jumps only at the design points t_i and, possibly, at infinity. Since the NPMLE solves a so–called generalized isotonic regression problem (see e.g. Robertson et al. 1988, Section 1.5), it coincides with the least squares estimator, i.e. it minimizes $q(\cdot)$ as well.

Now let us assume that the (sub-) distribution function F is concave on $[0, \infty)$. That means, it has a non-increasing density on $[0, \infty)$ and possibly a point mass at ∞. Then the LS estimator for the distribution function F is given by $\widehat{F}_{(3)}$ as in the previous section with $\boldsymbol{J} := [0, \infty)$, $I = 1$, $s_1 = 0$ and $[v_1, w_1] = [0, \infty]$. Here we assume without loss of generality that $0 \leq \widehat{F}_{(3)} \leq 1$, because all values Y_i are bounded from above by one, and $\min(G, 1) \in \mathcal{F}_{(3)}$ for any $G \in \mathcal{F}_{(3)}$.

Note also that Condition II is automatically satisfied with $\sigma^2 = 1/4$; see the proof of Hoeffding's (1963) inequality. Thus Conditions I and III together imply the conclusions of Theorem 1 and Corollary 1.

A final remark. As we just mentioned, without further constraints on F, the LS estimator and the NPMLE are identical. With the additional assumption of concavity, we may define the NPMLE $\widehat{F}_{ML}$ as the maximizer of $\ell(G)$ over the class of concave subdistribution functions on $[0, \infty)$. Characterizations, algorithms and consistency results for $\widehat{F}_{ML}$ are given by Dümbgen et al. (2003). We conjecture that the present results for the LS estimator hold for $\widehat{F}_{ML}$ as well. But the subsequent example shows that $\widehat{F}_{(3)} \neq \widehat{F}_{ML}$ in general.

A Counterexample. Suppose that $(t_1, Y_1) = (1, 0)$ and $(t_2, Y_2) = (2, 1)$. First consider the NPMLE, maximizing $\log(1 - G(1)) + \log G(2)$. Given a fixed value $G(2) = \alpha$, this function is maximized by taking $G(1)$ as small as possible under the constraints of concavity and $G(0) \geq 0$, so $G(1) = \alpha/2$. Since the function $\alpha \mapsto \log(1 - \alpha/2) + \log \alpha$ takes its maximum over $[0, 1]$ at $\alpha = 1$, we get that $\widehat{F}_{ML}(1) = 1/2$ and $\widehat{F}_{ML}(2) = 1$.

Now consider the LS estimator, minimizing $G(1)^2 + (1 - G(2))^2$. Again, for $G(2) = \alpha$ fixed, this function is minimized by $G(1) = \alpha/2$. Since the function $\alpha \mapsto (\alpha/2)^2 + (1 - \alpha)^2$ attains its minimum at $\alpha = 4/5$, we get that $\widehat{F}_{(3)}(1) = 2/5$ and $\widehat{F}_{(3)}(2) = 4/5$. Hence, $\widehat{F}_{(2)} \neq \widehat{F}_{ML}$.

5 Proofs

Our proof of Theorem 1 is based on directional derivatives of the sum of squared residuals. Let $\Delta : \boldsymbol{R} \to \boldsymbol{R}$ such that $\widehat{F} + \lambda\Delta$ is concave on $\boldsymbol{J}$ for some $\lambda > 0$. Then the optimality of $\widehat{F}$ implies that

$$0 \ \leq \ \frac{d}{d\lambda}\Big|_{\lambda=0} \sum_{i=1}^n (Y_i - (\widehat{F} + \lambda\Delta)(t_i))^2 \ = \ 2\sum_{i=1}^n \Delta(t_i)(\widehat{F}(t_i) - Y_i),$$

which is equivalent to

$$-\sum_{i=1}^n \Delta(t_i)E_i \ \geq \ \sum_{i=1}^n \Delta(t_i)(F - \widehat{F})(t_i). \tag{3}$$

In what follows we apply (3) to a special class of perturbation functions Δ and write

$$\|\Delta\|_n \ := \ \Big(\sum_{i=1}^n \Delta(t_i)^2\Big)^{1/2}.$$

Lemma 1 *For an integer $m \geq 0$, let $\mathcal{D}_m$ be the family of all continuous, piecewise linear functions on $\boldsymbol{R}$ with at most m knots. Then for any fixed $\gamma > 4$,*

$$S_n(m) := \sup_{\Delta \in \mathcal{D}_m} \frac{\left|\sum_{i=1}^n \Delta(t_i)E_i\right|}{\|\Delta\|_n} \ \leq \ \gamma\,\sigma\,(m+1)^{1/2}(\log n)^{1/2} \quad \textit{for all } m \geq 0$$

with probability tending to one as $n \to \infty$.

Proof of Lemma 1. Condition II implies that

$$\mathbb{P}\Big\{\Big|\sum_{i=1}^n h(t_i)E_i\Big|\Big/\|h\|_n \geq \eta\Big\} \ \leq \ 2\exp(-\eta^2/(2\sigma^2)) \tag{4}$$

for any function h with $\|h\|_n > 0$ and arbitrary $\eta \geq 0$. This follows from standard arguments involving Markov's inequality. For $1 \leq j \leq k \leq n$, let

$$\phi_{jk}^{(1)}(t) \ := \ 1\{t \in [t_j, t_k]\}\,\frac{t - t_j}{t_k - t_j} \quad \text{and} \quad \phi_{jk}^{(2)}(t) \ := \ 1\{t \in [t_j, t_k]\}\,\frac{t_k - t}{t_k - t_j}$$

if $t_j < t_k$. Otherwise let $\phi_{jk}^{(1)}(t) := 1\{t = t_k\}$ and $\phi_{jk}^{(2)}(t) := 0$. This defines a collection Φ of at most n^2 different nonzero functions $\phi_{jk}^{(e)}$. Then (4) implies that for any fixed $\gamma_o > 2$,

$$S_n := \max_{\phi \in \Phi} \Big|\sum_{i=1}^n \phi(t_i)E_i\Big|\Big/\|\phi\|_n \ \leq \ \gamma_o\sigma(\log n)^{1/2} \tag{5}$$

with probability tending to one as $n \to \infty$. For let $G_n(\phi) := \|\phi\|_n^{-1}\sum_{i=1}^n \phi(t_i)E_i$. Then, by (4),

$$\begin{aligned}
\mathbb{P}\{S_n \geq \gamma_o\sigma(\log n)^{1/2}\} \ &\leq \ \sum_{\phi \in \Phi} \mathbb{P}\Big\{|G_n(\phi)| \geq \gamma_o\sigma(\log n)^{1/2}\Big\} \\
&\leq \ 2n^2\exp(-\gamma_o^2\log(n)/2) \\
&\to \ 0 \quad \text{as } n \to \infty.
\end{aligned}$$

Now for any $\Delta \in \mathcal{D}_m$, there are $m' \leq 2m+2$ disjoint intervals on which Δ is either linear and nonnegative, or linear and nonpositive. For one such interval B with $M_n(B) > 0$ let $\{t_1, \ldots, t_n\} \cap B = \{t_j, \ldots, t_k\}$. Then

$$\Delta(t) \;=\; \Delta(t_k)\phi_{jk}^{(1)}(t) + \Delta(t_j)\phi_{jk}^{(2)}(t) \quad \text{for } t \in [t_j, t_k].$$

This shows that there are real coefficients $\lambda_1, \ldots, \lambda_{4m+4}$ and functions $\phi_1, \ldots, \phi_{4m+4}$ in Φ such that $\Delta = \sum_{j=1}^{4m+4} \lambda_j \phi_j$ on $\{t_1, \ldots, t_n\}$, and $\lambda_j \lambda_k \phi_j \phi_k \geq 0$ for all pairs (j, k). Consequently, inequality (5) entails that

$$\begin{aligned}
\frac{\left|\sum_{i=1}^n \Delta(t_i) E_i\right|}{\|\Delta\|_n} \;&\leq\; \frac{\sum_{j=1}^{4m+4} |\lambda_j| \left|\sum_{i=1}^n \phi_j(t_i) E_i\right|}{\left(\sum_{j=1}^{4m+4} \lambda_j^2 \|\phi_j\|_n^2\right)^{1/2}} \\
&\leq\; \frac{\sum_{j=1}^{4m+4} |\lambda_j| \|\phi_j\|_n}{\left(\sum_{j=1}^{4m+4} \lambda_j^2 \|\phi_j\|_n^2\right)^{1/2}} S_n \\
&\leq\; (4m+4)^{1/2} S_n \\
&\leq\; 2\gamma_o (m+1)^{1/2} \sigma (\log n)^{1/2},
\end{aligned}$$

by the Cauchy-Schwarz inequality. □

The next ingredient for our proof is a claim about differences of concave functions, which is similar to Lemma 5.2 of Dümbgen (1998); for the reader's convenience a proof will be given here.

Lemma 2 *Suppose that F satisfies Condition III. There is a universal constant $K = K(\beta, L) > 0$ with the following property: For any $\epsilon > 0$, let $\delta := K \min(b - a, \epsilon^{1/\beta})$. Then*

$$\sup_{t \in [a,b]} (\widehat{F} - F)(t) \;\geq\; \epsilon \quad \textit{or} \quad \sup_{t \in [a+\delta, b-\delta]} (F - \widehat{F})(t) \;\geq\; \epsilon$$

implies that

$$\inf_{t \in [c, c+\delta]} (\widehat{F} - F)(t) \;\geq\; \epsilon/4 \quad \textit{or} \quad \inf_{t \in [c, c+\delta]} (F - \widehat{F})(t) \;\geq\; \epsilon/4$$

for some $c \in [a, b-\delta]$.

Proof of Lemma 2. Suppose that $(\widehat{F} - F)(t_o) \geq \epsilon$ for some $t_o \in [a, b]$. Without loss of generality let $t_o \leq (a+b)/2$. We define an auxiliary linear function $\widetilde{F}$ via

$$\widetilde{F}(t) \;:=\; \begin{cases} F(t_o) & \text{if } \beta = 1 \\ F(t_o) + F'(t_o)(t - t_o) & \text{if } \beta > 1 \end{cases}$$

and note that

$$|(\widetilde{F} - F)(t)| \ \le \ L|t - t_o|^\beta/\beta, \tag{6}$$

by Condition III. Now let $0 < \delta \le (b-a)/8$. Since $\widehat{F} - \widetilde{F}$ is concave, it follows from $(\widehat{F} - \widetilde{F})(t_o + \delta) \ge \epsilon/2$ that $\widehat{F} - \widetilde{F} \ge \epsilon/2$ on $[t_o, t_o + \delta]$. Otherwise, if $(\widehat{F} - \widetilde{F})(t_o + \delta) < \epsilon/2$, then the derivative of $\widehat{F} - \widetilde{F}$ is less than or equal to $-\delta^{-1}\epsilon/2$ on $[t_o + \delta, \infty)$. Consequently, for $t \ge t_o + 3\delta$,

$$(\widehat{F} - \widetilde{F})(t) \ \le \ \epsilon/2 - \delta^{-1}(\epsilon/2)(t - t_o - \delta) \ \le \ -\epsilon/2.$$

Thus $\widehat{F} - \widetilde{F} \ge \epsilon/2$ or $\widetilde{F} - \widehat{F} \le \epsilon/2$ on some interval $J \subset [t_o, t_o + 4\delta]$ with length δ. Together with (6) this entails that $\widehat{F} - F$ or $F - \widehat{F}$ is not smaller than $\epsilon/2 - L(4\delta)^\beta/\beta \ge \epsilon/4$ on J, provided that $\delta \le (\beta/L)^{1/\beta}4^{-1-1/\beta}\epsilon^{1/\beta}$.

Now suppose that $(F - \widehat{F})(t_o) \ge \epsilon$ for some $\epsilon > 0$ and $t_o \in [a + \delta, b - \delta]$, where $0 < \delta \le (b-a)/2$. By Condition III and concavity of $\widehat{F}$ there exist numbers $\gamma, \widehat{\gamma}$ such that

$$\begin{aligned} F(t) \quad &\ge \quad F(t_o) + \gamma(t - t_o) - L|t - t_o|^\beta/\beta, \\ \widehat{F}(t) \quad &\le \quad \widehat{F}(t_o) + \widehat{\gamma}(t - t_o). \end{aligned}$$

Thus

$$(F - \widehat{F})(t) \ \ge \ \epsilon + (\gamma - \widehat{\gamma})(t - t_o) - L|t - t_o|^\beta/\beta \ \ge \ \epsilon - L\delta^\beta$$

for all t in the interval $[t_o, t_o + \delta]$ or $[t_o - \delta, t_o]$, depending on the sign of $\gamma - \widehat{\gamma}$. Moreover, $\epsilon - L\delta^\beta/\beta \ge \epsilon/4$, provided that $\delta \le (3/4)^{1/\beta}(\beta/L)^{1/\beta}\epsilon^{1/\beta}$. $\Box$

Finally we have to show that one of our classes $\mathcal{D}_m$ does indeed contain useful perturbation functions Δ. For that purpose we define the set

$$\mathcal{T} \ := \ \{t_1, t_2, \ldots, t_n\}$$

and denote with $\check{F}$ the unique continuous and piecewise linear function with knots in $\mathcal{T} \cap (t_1, t_n)$ such that $\check{F} = \widehat{F}$ on $\mathcal{T}$. Thus $\check{F}$ is one particular LS estimator for F.

Lemma 3 *For* $0 < u \le b - a$ *let*

$$\widetilde{M}_n(u) \ := \ \min_{c \in [a, b-u]} M_n[c, c+u].$$

Suppose that $F - \widehat{F} \ge \epsilon > 0$ or $\widehat{F} - F \ge \epsilon$ on some interval $[c, c+\delta] \subset [a, b]$ with length $\delta > 0$. Then there is a function $\Delta \in \mathcal{D}_6$ such that

$$\check{F} + \lambda\Delta \text{ is concave for some } \lambda > 0, \tag{7}$$

$$\Delta(F - \widehat{F}) \ \ge\ \epsilon\Delta^2 \quad \text{on } \mathcal{T}, \tag{8}$$

$$\|\Delta\|_n^2 \ \ge\ n\widetilde{M}_n(\delta/2)/4. \tag{9}$$

Proof of Lemma 3. Without loss of generality we assume that $\mathcal{T} \cap [c, c+\delta] \neq \emptyset$. For otherwise, $\widetilde{M}_n(\delta) \le M_n[c, c+\delta] = 0$, so that $\Delta \equiv 0$ would satisfy (7–9).

We define the auxiliary set

$$\mathcal{S} \ :=\ \Big\{t \in \boldsymbol{R} : \check{F}'(t-) > \check{F}'(t+)\Big\} \ \subset\ \mathcal{T} \cap (t_1, t_n).$$

Then for any function $\Delta \in \mathcal{D}_6$, requirement (7) is equivalent to

$$\Big\{t \in \boldsymbol{R} : \Delta'(t-) < \Delta'(t+)\Big\} \ \subset\ \mathcal{S}. \tag{10}$$

Case I: $\widehat{F} - F \ge \epsilon$ on $[c, c+\delta]$. Here a function $\Delta \in \mathcal{D}_4$ will do.

Case Ia: $\mathcal{S} \cap (c, c+\delta)$ contains some point t_o. We take $\Delta \in \mathcal{D}_3$ with knots $c, t_o, c+\delta$, where $\Delta = 0$ on $(-\infty, c] \cup [c+\delta, \infty)$ and $\Delta(t_o) = -1$. This function Δ satisfies (10) and (8). Moreover, $\Delta^2 \ge 1/4$ on some subinterval of $[c, c+\delta]$ with length $\delta/2$, whence $\|\Delta\|_n^2 \ge n\widetilde{M}_n(\delta/2)/4$.

Case Ib: $\mathcal{S} \cap (c, c+\delta) = \emptyset$. Now $F - \check{F}$ is concave on $[c, c+\delta]$, and $F - \check{F} \le -\epsilon$ on $[c, c+\delta] \cap \mathcal{T}$. Let $[c_o, d_o] \supset [c, c+\delta]$ be a maximal interval in $[-\infty, \infty]$ such that $F - \check{F}$ is concave on $[c_o, d_o] \cap \mathcal{T}$. Note that $c_o \in \mathcal{S}$ if $c_o > -\infty$, and $d_o \in \mathcal{S}$ if $d_o < \infty$. One easily verifies that there exists a linear function $\widetilde{\Delta}$ such that $\widetilde{\Delta} \ge F - \check{F}$ on $[c_o, d_o] \cap \mathcal{T}$ and $\widetilde{\Delta} \le -\epsilon$ on $[c, c+\delta] \cap \mathcal{T}$. Next let $(c_1, d_1) := \{\widetilde{\Delta} < 0\} \cap (c_o, d_o)$. Further let

$$c_2 \ :=\ \begin{cases} \max(\mathcal{T} \cap (-\infty, c_1)) & \text{if } c_1 > -\infty \text{ and } \widetilde{\Delta}(c_1) < 0, \\ \qquad c_1 & \text{else,} \end{cases}$$

$$d_2 \ :=\ \begin{cases} \min(\mathcal{T} \cap (d_1, \infty)) & \text{if } d_1 < \infty \text{ and } \widetilde{\Delta}(d_1) < 0, \\ \qquad d_1 & \text{else.} \end{cases}$$

Note that neither (c_2, c_1) nor (d_1, d_2) contains a design point t_i. Now let $\Delta \in \mathcal{D}_4$ with knots in $\{c_2, c_1, d_1, d_2\} \cap \boldsymbol{R}$ such that $\Delta = \widetilde{\Delta}/\epsilon$ on (c_1, d_1) and $\Delta = 0$ on $(-\infty, c_2) \cup (d_2, \infty)$. This function Δ satisfies (10) and (8). Moreover, $\Delta^2 \ge 1$ on $[c, c+\delta] \cap \mathcal{T}$, so that even $\|\Delta\|_n^2 \ge n\widetilde{M}_n(\delta)$.

Figure 1 illustrates the latter construction. For simplicity we only consider $\widehat{F} = \check{F}$. The upper subplot shows the graphs of F (thin line) and $\widehat{F}$ (thick line), the interval $[c, c+\delta]$ on which $\widehat{F} - F \geq \epsilon$ as well as the auxiliary points c_o, d_o. The lower subplot shows the corresponding perturbation function Δ (thick line) and the scaled difference $(F - \widehat{F})/\epsilon$ (thin line).

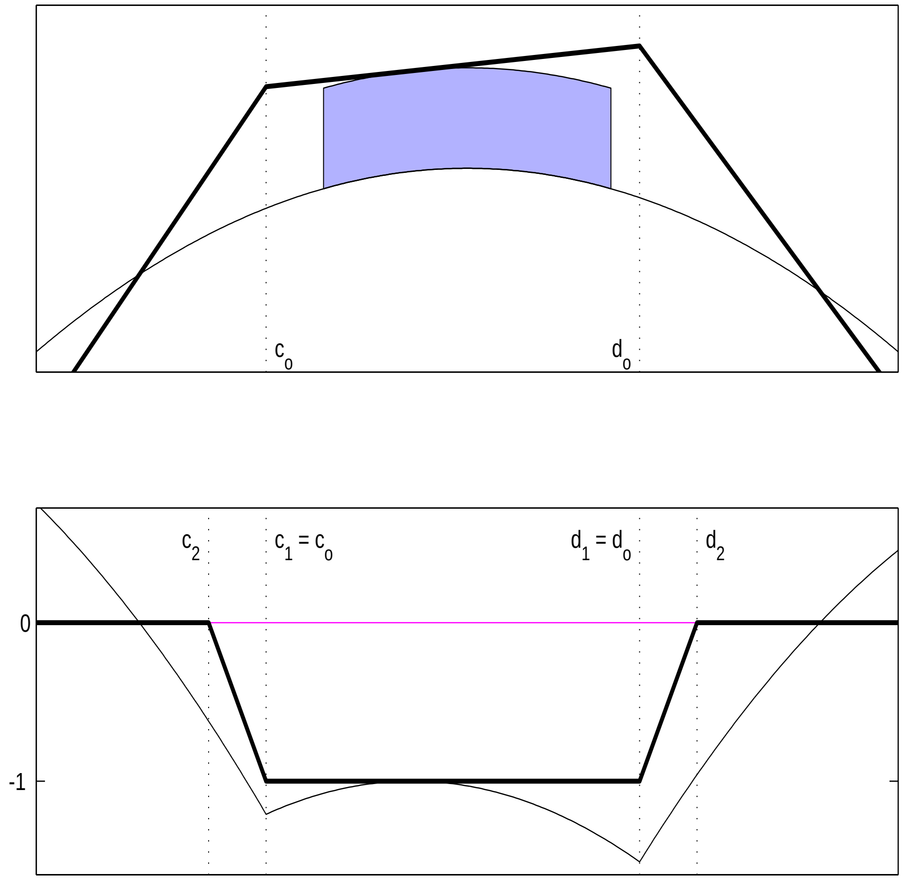

Figure 1: The perturbation function Δ in Case Ib

Case II: $F - \widehat{F} \geq \epsilon$ *on* $[c, c + \delta]$. Let $[c_o, d_o] \supset [c, c + \delta]$ be a maximal interval in $[-\infty, \infty]$ such that $F - \check{F} \geq \epsilon$ on $[c_o, d_o] \cap \mathcal{T}$. Now we define a function $\Delta \in \mathcal{D}_6$ as follows. At first let $\Delta := 1$ on $[c_o, d_o]$. Suppose that $d_o < \infty$, which implies that $d_o < t_n$. Then let d_1 be the largest number in $(d_o, \infty]$ such that $\check{F}$ is linear on $[d_o, d_1)$. Note that $F - \check{F}$ is concave on $[d_o, d_1] \cap \boldsymbol{R}$ and strictly decreasing on $[d_o, d_1] \cap \mathcal{T}$, where we define $F := -\infty$ on $\boldsymbol{R} \setminus \boldsymbol{J}$. Next let Δ be linear on $[d_o, d_1] \cap \boldsymbol{R}$ such that $\Delta(d_o) = 1$ and $\Delta(t_o) = 0$. Here t_o is the supremum of all points $t \in [d_o, d_1] \cap \boldsymbol{R}$ with $(F - \check{F})(t) \geq 0$.

If d_1 is finite, then it belongs necessarily to $\mathcal{S}$, and we define

$$d_2 := \begin{cases} d_1 & \text{if } t_o = d_1, \\ \min(\mathcal{T} \cap (d_1, \infty)) & \text{else.} \end{cases}$$

Then let $\Delta := 0$ on $[d_2, \infty)$, and let Δ be linear on $[d_1, d_2]$.

With an analogous construction in case of $c_o > -\infty$ we end up with a function $\Delta \in \mathcal{D}_6$ satisfying (10) and (8), while $\|\Delta\|_n^2 \geq n M_n[c, c+\delta] \geq n\widetilde{M}_n(\delta)$.

Figure 2 illustrates the latter construction. The upper subplot shows the graphs of F (thin line) and $\widehat{F} = \check{F}$ (thick line), the interval $[c_o, d_o]$ on which $F - \widehat{F} \geq \epsilon$ as well as the auxiliary points c_2, c_1, d_1, d_2. The lower subplot shows the corresponding perturbation function Δ (thick line) as well as $(F - \widehat{F})/\epsilon$ (thin line). $\square$

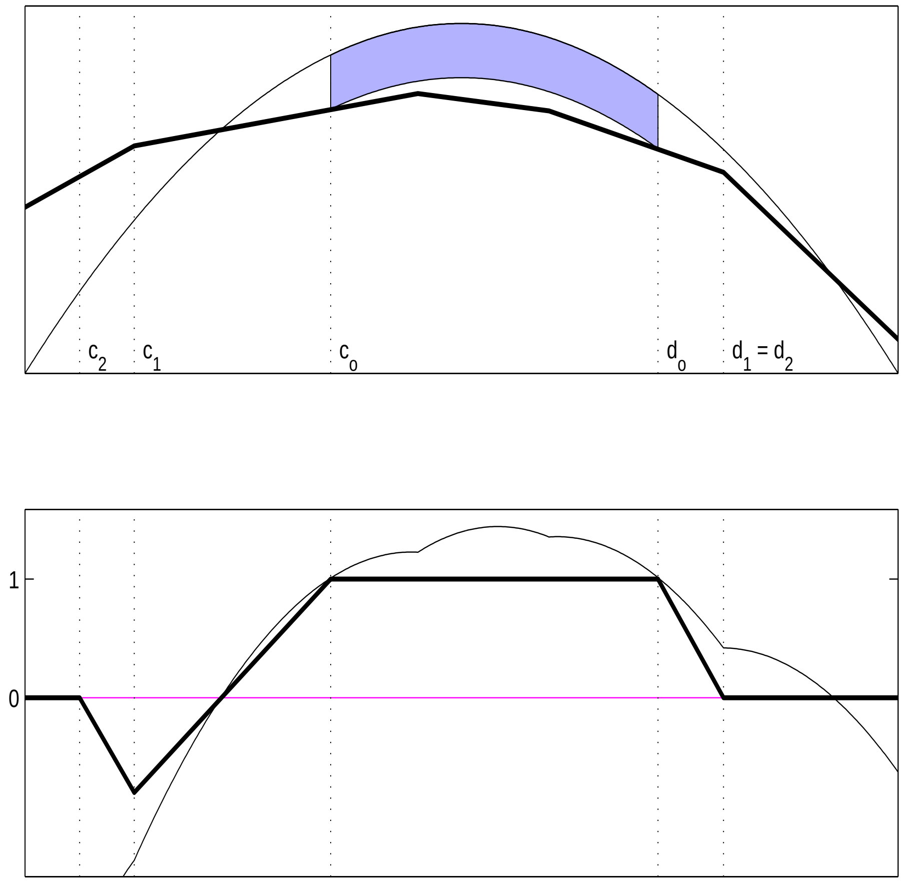

Figure 2: The perturbation function Δ in Case II

Proof of Theorem 1. Suppose that

$$\sup_{t \in [a,b]} (\widehat{F} - F)(t) \ \geq\ \kappa \delta_n^\beta \quad \text{or} \quad \sup_{t \in [a+\delta_n, b-\delta_n]} (F - \widehat{F})(t) \ \geq\ \kappa \delta_n^\beta$$

for some $\kappa > 0$. It follows from Lemma 2 that there is a (random) interval $[c_n, c_n + \delta_n] \subset [a, b]$ on which either $\widehat{F} - F \geq (\kappa/4)\delta_n^\beta$ or $F - \widehat{F} \geq (\kappa/4)\delta_n^\beta$, provided that n is sufficiently large and $\kappa \geq K^{-\beta}$. But then, by the definition of $S_n(6)$ and Lemma 3, there is a (random) function $\Delta_n \in \mathcal{D}_6$ such that

$$\begin{aligned}
S_n(6) \quad &\geq \quad -\|\Delta_n\|_n^{-1} \sum_{i=1}^n \Delta_n(t_i) E_i \\
&\overset{(3,7)}{\geq} \quad \|\Delta_n\|_n^{-1} \sum_{i=1}^n \Delta_n(t_i)(F - \widehat{F})(t_i) \\
&\overset{(8)}{\geq} \quad (\kappa/4)\delta_n^\beta \|\Delta_n\|_n \\
&\overset{(9)}{\geq} \quad (\kappa/4)\delta_n^\beta (n\widetilde{M}_n(\delta_n/2)/4)^{1/2}.
\end{aligned}$$

Consequently, by Condition I and Lemma 1,

$$\begin{aligned}
\kappa \quad &\leq \quad 4\delta_n^{-\beta}(n\widetilde{M}_n(\delta_n/2)/4)^{-1/2} S_n(6) \\
&\leq \quad 4\delta_n^{-\beta}\Big((C/8 + o(1))n\delta_n\Big)^{-1/2} S_n(6) \\
&= \quad O(1)(\log n)^{-1/2} S_n(6) \\
&= \quad O_p(1). \qquad \Box
\end{aligned}$$

Proof of Corollary 1. Let

$$\max_{t \in [a+\delta_n/2, b-\delta_n/2]} |\widehat{F}(t) - F(t)| \;=\; \delta_n^\beta R_n.$$

The proof of Theorem 1 reveals that $R_n = O_p(1)$. By concavity of $\widehat{F}$, for any $t \in [a + \delta_n, b - \delta_n]$ and $\nu_n := \delta_n/2$,

$$\frac{\widehat{F}(t) - \widehat{F}(t - \nu_n)}{\nu_n} \;\geq\; \widehat{F}'(t\,-) \;\geq\; \widehat{F}'(t\,+) \;\geq\; \frac{\widehat{F}(t + \nu_n) - \widehat{F}(t)}{\nu_n},$$

where $\widehat{F}'(t\,-)$ and $\widehat{F}'(t\,+)$ denote the left- and rightsided derivative of $\widehat{F}$, respectively. Moreover, the definition of R_n, concavity of F and Condition III with $\beta > 1$ imply that

$$\begin{aligned}
\frac{\widehat{F}(t) - \widehat{F}(t - \nu_n)}{\nu_n} \quad &\leq \quad \frac{F(t) - F(t - \nu_n) + 2\delta_n^\beta R_n}{\nu_n} \\
&\leq \quad F'(t - \nu_n) + 2\delta_n^\beta R_n/\nu_n \\
&\leq \quad F'(t) + L\nu_n^{\beta-1} + 2\delta_n^\beta R_n/\nu_n \\
&= \quad F'(t) + (2^{1-\beta}L + 4R_n)\rho_n^{(\beta-1)/(2\beta+1)}.
\end{aligned}$$

Similarly,

$$\frac{\widehat{F}(t+\nu_n)-\widehat{F}(t)}{\nu_n} \geq F'(t) - (2^{1-\beta}L + 4R_n)\rho_n^{(\beta-1)/(2\beta+1)}.$$

Hence we obtain

$$|\widehat{F}'(t\pm) - F'(t)| \ \leq \ (2^{1-\beta}L + 4R_n)\rho_n^{(\beta-1)/(2\beta+1)} \ = \ O_p\left(\rho_n^{(\beta-1)/(2\beta+1)}\right). \qquad \Box$$

Proof of Theorem 2. A close inspection of the proof of Theorem 1 reveals that we only need a surrogate for Lemma 3. Namely let $\mathcal{T}_{(1)} := \mathcal{T}$ and $\mathcal{T}_{(2)} := \mathcal{T}_{(3)} := \mathcal{T} \cup \{s_1, \ldots, s_I\}$. Let $\check{F}_{(j)}$ be the unique continuous and piecewise linear function with knots in $\mathcal{T}_{(j)} \cap \big(\min(\mathcal{T}_{(j)}), \max(\mathcal{T}_{(j)})\big)$ such that $\check{F}_{(j)} = \widehat{F}_{(j)}$ on $\mathcal{T}_{(j)}$. Then we have to show that Lemma 3 remains true with $(\widehat{F}_{(j)}, \check{F}_{(j)}, \mathcal{T}_{(j)})$ in place of $(\widehat{F}, \check{F}, \mathcal{T})$, where Condition (7) has to be replaced with

$$\check{F}_{(j)} + \lambda\Delta \ \in \ \mathcal{F}_{(j)} \quad \text{for some } \lambda > 0. \tag{11}$$

For that purpose we use the same construction of Δ as in the proof of Lemma 3. In order to guarantee (11), since $\check{F}_{(j)}$ and Δ are piecewise linear with $\check{F}_{(j)} \in \mathcal{F}_{(j)}$, it suffices to verify the following two conditions:

$$\text{If } j \in \{1,3\}, \text{ then} \quad \Delta'(t+) \geq 0 \text{ whenever } \check{F}'_{(j)}(t+) = 0. \tag{12}$$

$$\text{If } j \in \{2,3\}, \text{ then for } 1 \leq i \leq I, \tag{13}$$

$$\Delta(s_i) \begin{cases} \geq & 0 \quad \text{if } \check{F}_{(j)}(s_i) = v_i, \\ \leq & 0 \quad \text{if } \check{F}_{(j)}(s_i) = w_i. \end{cases} \tag{14}$$

Let us start with (12), where $j \in \{1,3\}$. Note first that $\check{F}'_{(j)}(t+) > 0$ for all $t < \max(\mathcal{S})$, where $\mathcal{S} = \{t : \check{F}'_{(j)}(t-) > \check{F}'_{(j)}(t+)\}$. Now we consider Case Ia as defined in the proof of Lemma 3. There the function Δ satisfies $\Delta'(\cdot+) \geq 0$ on $[t_o, \infty)$, for some $t_o \in \mathcal{S} \cap (c, c+\delta)$, which entails (12).

In Case Ib, if $d_1 < \infty$ and $\widetilde{\Delta}(d_1) < 0$, then $d_1 = d_o$ belongs to $\mathcal{S}$ and $\Delta'(t+) \geq 0$ for all $t \geq d_1$. If $d_1 < \infty$ and $\widetilde{\Delta}(d_1) = 0$, then $\Delta'(t+) \geq 0$ for all $t \geq c_1$, and $c_1 > -\infty$ entails that $c_1 \in \mathcal{S}$. Thus (12) is satisfied in case of $d_1 < \infty$. If $d_1 = \infty$ and $\widetilde{\Delta}' < 0$, suppose that $\widetilde{\Delta}' < 0$ was really necessary. That means, there exists a point $r \in (c_o, c) \cap \mathcal{T}_{(j)}$ with $(F - \check{F}_{(j)})(r) > -\epsilon$. Since $(F - \check{F}_{(j)})(s) \leq -\epsilon$ for some $s \geq c$, this implies that $F'(t+) - \check{F}'_{(j)}(t+) < 0$ for all $t \geq s$. But $F'(\cdot+) \geq 0$, so that $\check{F}_{(j)}(\cdot+) > 0$ everywhere, whence (12) is trivial.

Now consider Case II. If $\Delta'(t+) < 0$ for some $t < c_o$, then t is strictly smaller than some point within $\mathcal{S}$, by construction of Δ. If $\Delta'(t+) < 0$ for some $t \geq d_o$, then either $t < d_1 \in \mathcal{S}$, or $\check{F}_{(j)}$ is linear on $[d_o, \infty)$. In the latter case, there exists a point $s \in \mathcal{T}_{(j)} \cap (d_o, \infty)$ such that $(F - \check{F}_{(j)})(d_o) \geq \epsilon > (F - \check{F}_{(j)})(s)$. This entails that $(F - \check{F}_{(j)})'(s+) < 0$. Hence $\check{F}'_{(j)}(s+) > F'(s+) \geq 0$, so that $\check{F}'_{(j)}(\cdot+) > 0$ everywhere.

As for Condition (13), note that $\Delta(F - \check{F}_{(j)}) \geq \epsilon\Delta^2$ on $\mathcal{T}_{(j)}$. In particular, if $j \in \{2, 3\}$, then $\Delta(s_i) < 0$ implies that $F(s_i) - \check{F}_{(j)}(s_i) < 0$, i.e. $\check{F}_{(j)}(s_i) > F(s_i) \geq v_i$. Similarly, $\Delta(s_i) > 0$ entails that $\check{F}_{(j)}(s_i) < w_i$. $\square$

Acknowledgement. We are grateful to a referee for constructive comments.

References

DÜMBGEN, L. (1998). New goodness-of-fit tests and their application to nonparametric confidence sets. *Ann. Statist.* ***26***, 288–314

DÜMBGEN, L., FREITAG, S., AND JONGBLOED, G. (2003). Estimating a unimodal distribution from interval-censored data. Preprint

GROENEBOOM, P., JONGBLOED, G., AND WELLNER, J.A. (2001). Estimation of a convex function: characterization and asymptotic theory. *Ann. Statist.* ***29***, 1653–1698

GROENEBOOM, P., AND WELLNER, J.A. (1992). *Information bounds and nonparametric maximum likelihood estimation.* Birkhäuser, Basel.

HANSON D.L. AND PLEDGER, G. (1976). Consistency in Concave Regression. *Ann. Statist.* ***4***, 1038–1050

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* ***58***, 13–30

IBRAGIMOV, I.A. AND R.Z. KHASMINSKII (1980). Estimates of signal, its derivatives, and point of maximum for Gaussian observations. *Theory Prob. Appl.* ***25***, 703-716.

MAMMEN, E. (1991). Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.* ***19***, 741–759

ROBERTSON, T., WRIGHT, F.W. AND DYKSTRA, R.L. (1988). *Order restricted statistical inference.* Wiley, New York.

VAN DE GEER, S.A. (2000). *Empirical Processes in M-Estimation.* Cambridge University Press, Cambridge.

Lutz Dümbgen
Department of Mathematics
University of Bern
Sidlerstrasse 5
CH-3012 Bern, Switzerland
duembgen@stat.unibe.ch

Sandra Freitag
Institut für Med. Informatik und Statistik
Universität Kiel
Brunswiker Strasse 10
D-24105 Kiel, Germany
freitag@medinfo.uni-kiel.de

Geurt Jongbloed
Department of Mathematics
Vrije Universiteit
De Boelelaan 1081A
1081 HV Amsterdam, The Netherlands
geurt@cs.vu.nl